U-net combined with CRF and anatomical based spatial features to segment white matter hyperintensities

PengZheng Zhou[1,2], Li Liang[1,2], XuTao Guo[1,2], Haiyan Lv[3], Tong Wang[1], *Member IEEE,*
Ting Ma[1,2,4,5]*, *Member IEEE*

***Abstract*—White matter hyperintensities (WMH) are important biomarkers for cerebral small vessel disease and closely associated with other neurodegenerative process. In this paper, we proposed a fully automatic WMH segmentation method based on U-net architecture. CRF were combined with U-net to refine segmentation results. We used a new anatomical based spatial feature produced by brain tissue segmentation based on T1 image, along with intensities of T1 and T2-FLAIR images to train our neural network. We compared 8 forms of automated WMH segmentation methods, range from traditional statistical learnng methods to deep learning based methods, with different architecture and used different features. Results showed our proposed method achieved best performance in terms of most metrics, and the inclusion of anatomical based spatial features strongly increase the segmentation performance.**

I. Introduction

White matter hyperintensities (WMH) are brain areas in the cerebral white matter with increased signal intensity on T2-weighted or fluid-attenuated inversion recovery (FLAIR) magnetic resonance imaging (MRI) scans. WMH are commonly found in the brain of healthy elders and patients diagnosed with small vessel disease and other neurological disorders [1]. The location and size of WMH regions are important biomarkers of these diseases. Manual delineation by experienced neurologists is a reliable way to segment the WMH regions but it is expensive, time consuming, subjective and impractical for large scale longitudinal studies. As such, automatically segmenting the WMH regions is of great importance, especially in the context of large-scale neuroimaging studies. However, none of the methods achieved reliable performance close to human readers. Major challenges for WMH segmentation include: noise, imaging artifacts, inhomogeneous intensities, random locations, variability in size and shape, et al . The presence of other brain anatomies structures or pathologies with enhanced intensities on FLAIR image add further difficulties .

Fully automated WMH detection algorithms had been developed during the past decade. The FSL group had published a toolbox for WMH segmentation called BIANCA (Brain Intensity AbNormality Classification) [2]. Another public available toolbox called the Lesion Segmentation Toolbox (LST), as implemented in SPM12, included lesion prediction algorithm (LPA) [3] and lesion growth algorithm (LGA) [4]. More recently deep neural networks have achieved great success in both natural and medical image domains [5,6]. Deep learning based techniques have been successfully employed to segment WMH regions [7,8]. Some of these methods use single FLAIR contrast, while others employed multiple modalities including T1, T2, proton density, and even diffusion tensor imaging data.

In this paper, we present a customized U-Net with conditional random fields (CRF), namely U-Net+CRF, to segment WMH. The features fed into our network were: intensities of T1 and T2-FLAIR, combined with anatomical based spatial features obtained from T1 brain tissue segmentation. We compared 8 forms of WMH segmentation methods (3 traditional statistical learning methods: k-nearest neighbor, LPA and LGA, 5 deep learning based methods with two different network architectures: DeepMedic [9] and U-net [6]), results showed that the incorporate of anatomical based spatial features effectively improve the segmentation performance, and proposed U-Net+CRF with the inclusion of anatomical based spatial features achieved the best performance in terms of most metrics.

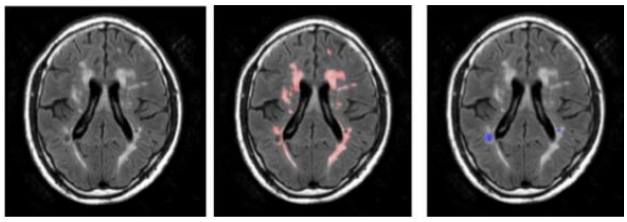
II. Method

A. Dataset

The dataset used in this study came from recently proposed WMH segmentation challenge in in Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2017 (https://wmh.isi.uu.nl/). Public available training set consists of 60 subjects acquired from three scanners. For each subject, T2-FLAIR and T1 MR images along with ground truth images of WMH regions were provided. Training data was annotated manually according to standards for reporting vascular changes on neuroimaging (STRIVE) [1]. There are three labels for each subject on T2-FLAIR MR imaging, 0 for background, 1 for WMH, and 2 for other pathology. as shown in Figure 1, the different lesions overlap on T2-FLAIR image.

This study was supported by grants from National Key Research and Development Program of China (2018YFC1312000), the National Natural Science Foundation of China (61801145), and the Basic Research Project of Shenzhen Science and Technology Program (JCYJ20180306171800589).

1. Department of Electronic and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, Guangdong, China.
2. Peng Cheng Laboratory, Shenzhen, Guangdong, China
3. Mindsgo Life Science Shenzhen Ltd, ShenZhen, Guangdong, China.
4. National Clinical Research Center for Geriatric Disorders, Xuanwu Hospital Capital Medical University, Beijing, China.
5. Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China.
*Corresponding author: Ting Ma. Phone: +86-755-26033608; fax: +86-755-26033608; email: tmahit@outlook.com



A) B) C)

Figure 1. An example of subject from WMH Segmentation Challenge of MICCAI 2017: A) T2-FLAIR image, B) WMH ROI overlap on T2-FLAIR image, C) other pathological ROI overlap on T2-FLAIR image.

B. Data pre-processing

Data preprocessing plays an import role in our overall framework. The flow chart of MR image pre-processing were showed on Figure 2. We used FMRIB software library (FSL) linear image registration tool (FLIRT) to co-register T1 and FLIAR images, with T2-FLAIR images served as reference space. Then we removed non brain tissue from T1 images using FSL brain extraction tool (BET), and resulting brain mask is applied to the FLAIR images. Radio frequency filed inhomogeneity correction is performed on both images using the N3 algorithms [10]. Intensity in MRI images varies among the patients. Therefore we normalized the intensity using the method proposed by Nyul et al [11]. We proposed anatomical based spatial feature information obtained from brain tissue segmentation of T1 images. T1 images were segmented into 283 regions of interest (ROIs) using a multi-atlas segmentation pipeline of Brain Label [12]. After coregistration, the T2-FLAIR images data were segmented into the same 283 ROIs. For the purpose of this study, we regrouped the finest level of ROIs into 26 white matter regions based on their ontological relationships [13]. We produced anatomical based spatial feature image for each subject by assigning a code to symmetric white matter regions, i.e. 13 codes were assigned to 13 symmetric white matter ROIs as shown on Table 1.

C. *Neural network architecture*

The network architecture, illustrated in Figure 3, is a 2D asymmetrical encoder-decoder network based on the U-Net [6] architecture. We used residual connections proposed by the 2D uResNet architecture [14] for WMH segmentation in MRI. We used residual blocks with convolution layers in the encoder and decoder. Instead of the more typical rectified linear unit (ReLU), we used in our residual blocks a parame-

Table 1. Label and codes for white matter ROIs

Brain region	Label	Normalized coding
Body of corpus callosum	1	0.06
Genu of corpus callosum	2	0.12
Splenium of corpus callosum	3	0.18
Anterior deep White Matter	4	0.24
Posterior deep White Matter	5	0.30
Anterior limb of internal capsule	6	0.36
Posterior limb of internal capsule	7	0.42
Limbic White Matter	8	0.48
Inferior deep White Matter	9	0.54
Frontal subcortical White Matter	10	0.60
Parietal subcortical White Matter	11	0.66
Temporal subcortical White Matter	12	0.72
Occipital subcortical White Matter	13	0.78
Cerebrospinal Fluid	14	0.01
Gray matter & Cerebellum	15	0.01

-tric version, the PReLU non-linearity. We perform down sampling in each resolution step by concatenating the result of a max pooling operation and strided convolution. This strategy avoids representational bottlenecks while keeping the number of parameters contained. Fully-connected Conditional Random Field (CRF) is often used as post-processing to refine voxel classification results by encouraging spatial coherence. In our neural network, we used the architecture of convolutional neural network that combine the strength of U-Net and CRF probabilistic graphical modelling. Importantly, our system fully integrates CRF modelling with U-Net, making it possible to train the whole deep network end-to-end with the usual back-propagation algorithm, avoiding offline post-processing methods for object delineation.

D. *Segmentation and post-processing*

Within this approach, the class imbalance issue is addressed with a combination of techniques including the use of small patches (48×48) and a weighted loss function. Data augmentation is an effective way to equip the deep networks with desired invariance and robustness properities when training data are limited.

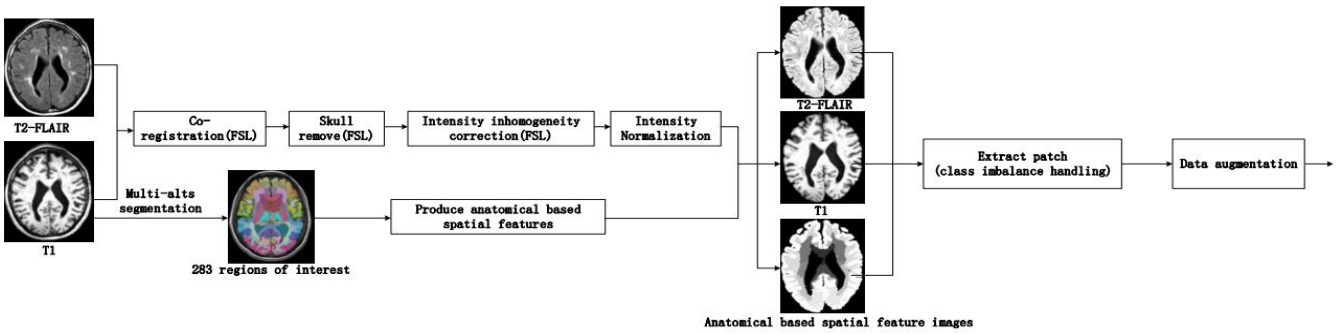


Figure 2. Basic process pipeline for data preprocessing. The extracted patches are of size 48×48 and include all input modalities.

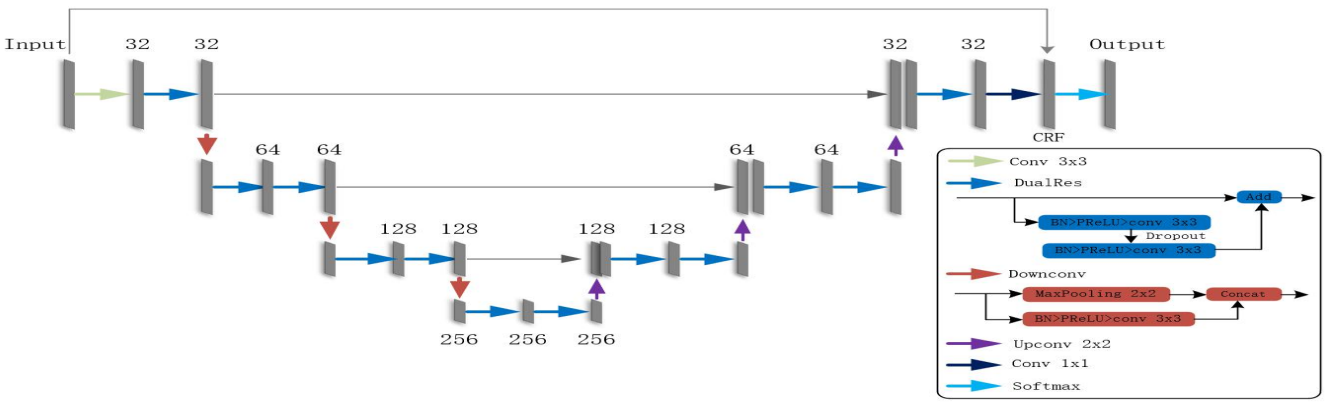


Figure 3. Neural network architecture

Once the network weights have been trained, we can segment the samples through the network, where the samples are extracted uniformly with size of 48 × 48 and the extraction step is 12. The resulting patch probabilities are then preserve their original spatial position to generate the entire volume probability map. In our case, combining is performed on each voxel by averaging the class probabilities of the individual patches. In addition, since the extraction step is smaller than the patch size, a certain degree of overlap is used between the extracted patches. Therefore, the same voxels are marked as visible in different neighborhoods and the resulting class probabilities are averaged. Finally, the probability maps are binarized by thresholding the probability of the lesion category and then performing connected component filtering by the lesion volume.

E. Evalution

Five different metrics evaluate the segmentation performance in different aspects as proposed in MICCAI 2017 WMH segmentation challenge [15]: (1) the Dice Similarity Coefficient (DSC), (2) a modified Hausdorff distance (95th percentile; H95), (3) the absolute percentage volum difference (AVD), (4) the sensitivity for detecting individual lesion (recall), (5) F1-score for individual lesion (F1).

III. Results

The proposed U-net+CRF method was compared to other well established and state-of-the-art algorithms. From the lesion segmentation toolbox (LST) LPA and LGA frame-works are used. For LPA we only used FLAIR images as input, while LGA used FLAIR and T1 images. The FSL group had published a toolbox for WMH segmentation called BIANCA. DeepMedic is a recently released CNN library for medical image segmentation. The default settings are used in the comparison performed here. All the methods were evaluated using a 5-fold cross validation. The dataset was split in such a way that all the 60 images are tested at least once. For each fold, 12 subjects were picked for test, the remaining 48 are used for training. For evaluation, DSC, H95, AVD, Recall, F1 are used. We used the keras to build our network and train it in NVIDIA Quadro GV100 GPU.

As shown in Table 2. For all five metrics, it can be observed that the neural network (DeepMedic, U-Net) perform better than the traditional methods (BIANCA, LGA, LPA). It is also evident that segmentation accuracy of the neural network can be improved by adding CRF (DeepMedic, U-Net+CRF). Especially, our proposed anatomical based spatial feature further improved the segmentation accuracy. The proposed U-Net+CRF+Spatial outperforms the rest methods in terms of most metrics.

We further evaluated the performance of our U-Net+CRF model for three sub groups. In our dataset, 11 subject were included in low WMH group (with total WMH volume 0 ml to 5 ml), 34 subject were included in median WMH group (with total WMH volume 6 ml to 15 ml), 15 subject were included in large WMH group (with total WMH volume > 15 ml). As show in Table 3, our U-Net+CRF+Spatial model outperforms the other methods for three groups in terms of DSC. Our model largely improves the accuracy of WMH detection in low and medium WMH group, which is preferable because WMH at early stage of several neurodegeneration diseases tend to have sparse distribution and small volume.

Table 2. Crossval-validation results for different WMH segmentation methods on the MICCAI 2017 dataset

Algorithm	DSC	H95(mm)	AVD	Recall	F1
BIANCA	0.54±0.23	14.78±9.54	0.26±0.25	0.58±0.18	0.21±0.12
LGA	0.57±0.21	14.28±10.73	0.24±0.29	0.15±0.09	0.23±0.12
LPA	0.64±0.19	9.47±7.77	0.21±0.23	0.41±0.21	0.40±0.14
DeepMedic	0.75±0.13	4.51±4.98	0.11±0.11	**0.80±0.10**	0.65±0.14
U-Net	0.69±0.22	8.55±8.13	0.19±0.28	0.68±0.14	0.59±0.20
U-Net+CRF	0.70±0.25	6.48±7.03	0.16±0.24	0.70±0.22	0.59±0.20
U-Net+Spatial	0.74±0.16	6.89±12.31	0.11±0.11	0.68±0.15	0.66±0.13
U-Net+CRF+Spatial	**0.78±0.11**	**3.70±4.49**	**0.10±0.09**	0.77±0.11	**0.67±0.12**

In figure 4, we showed the segmentation results of three representative subjects, The results were obtained from the U-Net+CRF+Spatial.

Table 3. Comparison of results for different methods of different total WMH load (S: low WMH group, M: medium WMH group and L: large WMH group)

Algorithm	DSC			H95 (mm)		
	S	M	L	S	M	L
BINANCA	0.28	0.53	0.73	24.95	16.25	6.93
LGA	0.39	0.52	0.72	24.43	13.83	6.95
LPA	0.48	0.63	0.76	17.09	9.47	3.87
Deepmedic	0.63	0.76	0.84	9.28	**2.46**	1.55
U-Net+CRF+Spatial	**0.68**	**0.77**	**0.86**	**7.35**	2.67	**1.30**

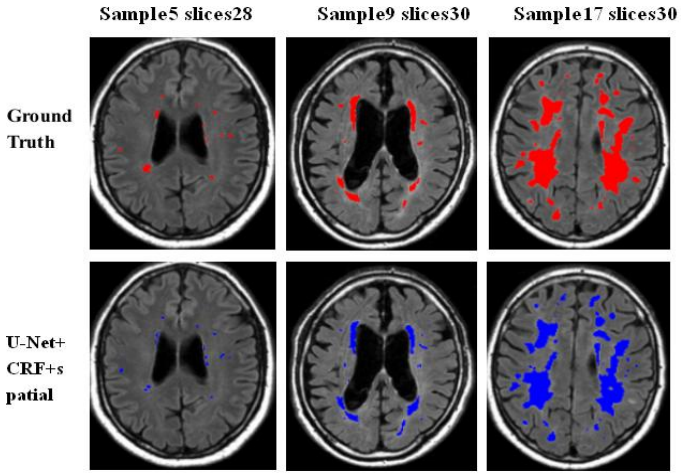


Figure 4 The segmentation of WMH used U-net+CRF+Spatial

IV. CONCLUSION

The main contributions of our study are: 1) We compared 8 forms of WMH segmentation methods, three of them are public available toolbox used traditional statistical learning strategy, 4 of them utilized U-net based architecture, one used network architecture of DeepMedic. We compared their performance of WMH segmentation with benchmark dataset. Results showed that LPA achieved best performance among three traditional statistical learning algorithms. But in general, deep learning based methods outperform LPA, LGA and BIANCA. 2) We proposed a U-net+CRF architecture to segment WMH, which achieved the best performance among all 8 methods in terms of most metrics. 3) We regrouped the results of brain tissue segmentation based on T1 image (with finest 283 brain regions) into 13 white matter regions and 2 non white matter regions to obtain anatomical based spatial feature image. The inclusion of anatomical based spatial feature largely improves the WMH segmentation performance.

ACKNOWLEDGMENT

This study was supported by grants from National Key Research and Development Program of China (2018YFC1312000), the National Natural Science Foundation of China (61801145), and the Basic Research Project of Shenzhen Science and Technology Program (JCYJ20180306171800589). Department of Electronic and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, Guangdong, China.

REFERENCES

[1] Jonna M Wardlaw, Eric E Smith, Greet J Biessels, et al. "Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration," The Lancet Neurology, vol.17, no. 1, pp. 1-18, 2013.

[2] Griffanti, L., Zamboni, G., Khan, A., et al, 2016. BIANCA (Brain intensity AbNormality Classification Algorithm): a new tool for automated segmentation of white matter hyperintensities. Neuroimage 141, 191–205.

[3] Schmidt, P., 2017. Bayesian Inference for Structured Additive Regression Models for Large-Scale Problems with Applications to Medical Imaging. Ludwig-Maximilians University, München.

[4] Schmidt, P., Gaser, C., Arsic, M., Buck, D., Forschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V.J., Zimmer, C., Hemmer, B., Muhlau, M., 2012. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. Neuroimage 59, 3774-3783.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Su "Deep residual learning for image recognition," in IEEE Conference on Computer Vision and Pattern Recognition, June 2016.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation,pp.234-241, Springer International Publis-hing, 2015.

[7] Ghafoorian, M., Karssemeijer, N., Heskes, T., Uder, I.W.M.v., Leeuw, F.E.d., Marchiori, E., Ginneken, B.v., Platel, B., 2016. Non-uniform patch sampling with deep convolutional neural networks for white matter hyperintensity segmentation. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1414-1417.

[8] Moeskops, P., de Bresser, J., Kuijf, H.J., Mendrik, A.M., Biessels, G.J., Pluim, J.P.W., Isgum, I., 2018. Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI. Neuroimage Clin. 17, 251-262.

[9] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis, 36:61–78, 2017.

[10] Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imageing1998; 17(1): 87-97.

[11] Nyul LG, Udupa JK. Standardizing the MR image intensity scales: making MR intensities have tissue-specific meaning. In: Medical imaging 2000. Interna-tional Society for Optics and Photonics; 2000. p. 496–504.

[12] BrainLabel: http://brainlabel.org.

[13] Wu D, Albert M, Soldan A, et al. Multi-atlas based detection and localization (MADL) for location-dependent quantification of white matter hyperintensities. Neuroimage Clin. 2019; 22: 101772.

[14] R. Guerrero, C. Qin, O. Oktay, C. Bowles, L. Chen, R. Joules, R. Wolz, M.C. Vald´es-Hern´andez, D.A. Dickie, J. Wardlaw, and D. Rueckert. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. NeuroImage: Clinical, 17:918 – 934, 2018. doi: 10.1016/J.NICL.2017.12.022.

[15] Hugo J. Kuijf, J. Matthijs Biesbroek, et al. Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities; Results of the WMH Segmentation Challenge. IEEE Trans Med Imaging. 2019 Mar 19.